



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2006

Rhythm and Speech Rate: A Variation Coefficient for deltaC

Dellwo, Volker

Abstract: The percentage of vocalic intervals (%V) and the standard deviation of consonantal intervals (deltaC) in a speech signal are two dimensions according to which languages of different rhythm classes (e.g. stress-timed, syllable-timed) seem to be differentiable on an acoustic level (Ramus et al., 1999). In this context it has been found that especially deltaC varies considerably as a function of speech rate (Barry et al., 2003 and Dellwo Wagner, 2003). The present paper argues that if deltaC was determined by speech rate it would describe speech rate rather than rhythm. For this reason a variation coefficient (varcoDeltaC) will be calculated in order to monitor relative deltaC variation across speech rates. Results for varcoDeltaC support the views that a) according to varcoDeltaC rhythm classes seem to be better differentiable and b) some languages tend to vary in rhythm as a function of speech rate (German, English), while the rhythm of other languages seems to be unaffected by changes in speech rate (French).

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-111789>

Book Section

Accepted Version

Originally published at:

Dellwo, Volker (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In: Karnowski, P; Szigeti, I. Language and language-processing. Frankfurt/Main: Peter Lang, 231-241.

Rhythm and Speech Rate: A variation coefficient for ΔC

Volker Dellwo

Department for Phonetics and Linguistics, University College London

Email: volker@phon.ucl.ac.uk

The percentage of vocalic intervals (%V) and the standard deviation of consonantal intervals (ΔC) in a speech signal are two dimensions according to which languages of different rhythm classes (e.g. stress-timed, syllable-timed) seem to be differentiable on an acoustic level (Ramus et al., 1999). In this context it has been found that especially ΔC varies considerably as a function of speech rate (Barry et al., 2003 and Dellwo & Wagner, 2003).

The present paper argues that if ΔC was determined by speech rate it would describe speech rate rather than rhythm. For this reason a variation coefficient ($\text{varco}\Delta C$) will be calculated in order to monitor relative ΔC variation across speech rates.

Results for $\text{varco}\Delta C$ support the views that a) according to $\text{varco}\Delta C$ rhythm classes seem to be better differentiable and b) some languages tend to vary in rhythm as a function of speech rate (German, English), while the rhythm of other languages seems to be unaffected by changes in speech rate (French).

1 Introduction

In their well known rhythm-class hypothesis Pike (1945) and Abercrombie (1967) argue that the languages of the world can be classified following two types of rhythm patterns: a) stress-timed rhythm and b) syllable-timed rhythm. According to this hypothesis both types of rhythm show rhythmical units of equal duration (isochrony hypothesis): stress-timed languages tend to have isochronous inter-stress intervals while syllable-timed languages tend to have rather equal isochronous syllable durations. Classic examples for a stress-timed language are English, Dutch, and German, while French, Spanish, and Italian have often been found to be syllable-timed¹.

The attribution of rhythm-classes to particular languages turned out to be based solely on intuitions since a vast amount of attempts to find acoustic correlates for isochrony in languages of the respective classes carried out mainly in the 1970s and 80s remained without success (cf. Grabe & Low, 2002, and Ramus et al., 1999, for detailed discussions). For this reason various researchers argued that syllable isochrony in syllable-timed and inter-stress isochrony in stress-timed languages is merely a perceptual phenomenon and is not represented in the durations of the respective intervals on an acoustic level (cf. Beckman, 1992, for a discussion).

Nevertheless, in the recent past new promising attempts for acoustic correlates of rhythm class in the speech signal have been proposed. These correlates are no longer based on the syllable or inter-stress interval as rhythmical units but on vocalic and intervocalic intervals, e.g. a measure based on the percentage of vocalic intervals (%V) and the standard deviation of consonantal intervals (ΔC) by Ramus et al. (1999) or the raw and normalised pairwise variability index (rPVI/nPVI) by Grabe & Low (2002) based on a pairwise comparison of the durations of either two consecutive vocalic or consonantal intervals.

¹ A third rhythm class, mora-timed languages as e.g. Japanese, has been identified, but it is not dealt with in the present study.

The present study deals exclusively with the proposal by Ramus et al. (1999) in which the authors find that stress-timed and syllable-timed languages cluster around different areas along the two dimensions %V and ΔC while stressed-timed languages show a higher ΔC and a lower %V than syllable-timed languages². The rationale behind this is that stress-timed languages (or at least the ones currently under observation) allow complex consonant clusters, thus have a higher ΔC , while this pattern is restricted for syllable-timed languages. A higher %V in syllable-timed languages is explained by the fact that these languages do not allow vowel reduction while this is a common feature of stress-timed languages (again this accounts for the languages currently under observation). Since %V and ΔC are temporal patterns in speech it has often been assumed that speech rate has a major impact on the two values (cf. Ramus, 2002, and Grabe & Low, 2002). In this respect Barry et al.

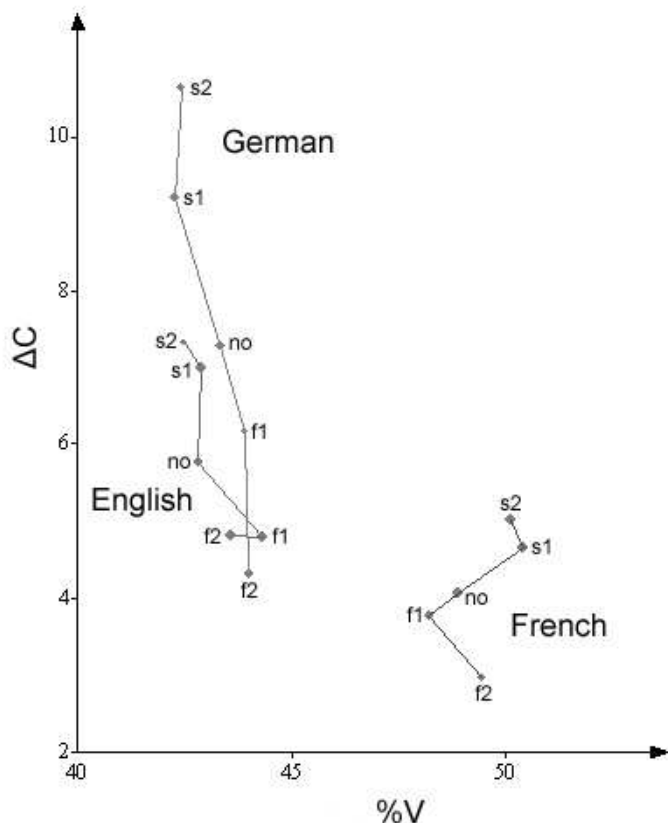


Diagram 1: Results for %V and ΔC under different intended speech rate conditions (s2, s1, no, f1, f2) for the languages English, French, and German.

(2003) and Dellwo & Wagner (2003) found that ΔC correlates negatively with speech rate (cf. diagram 1 for the results of the type found in Dellwo & Wagner, 2003). Additionally Barry et al. (2003) made this discovery for ΔV (the standard deviation of vocalic intervals). Thus it seems that standard deviations of either consonantal or vocalic intervals are to a great degree dependent on the overall speech rate at which a speaker performs.

The present study starts from the assumption already formulated in Dellwo & Wagner (2003) that this finding may have a simple rational explanation: Since in fast speech consonantal intervals may likely to be shorter than in slow speech this may have a direct effect on the extent to which the durations can vary on an absolute level: shorter intervals cause lower absolute variation, longer intervals cause higher absolute variation. Since ΔC is the absolute standard deviation of consonantal intervals lower values may be expected in fast speech than in slow speech.

This seems not only to be the case in speech; other examples from the 'real

world' show how trivial these observations may be: assume two people, the first having a total amount of 1,000 € on his/her bank account, the second 1,000,000 €. It is rather unlikely that the absolute amounts of monthly money transactions of person 1 are the same as for person 2. Since person 2 owns much more money it is able to spend more and rather likely to earn more as well in absolute terms than person 1 thus the standard deviation of monthly transactions will be higher for person 2.

Back to ΔC : In order to compare the variation between different speech rates it is important to compare relative variation to the norm rather than absolute variation (as has been done so far). A variation coefficient (varco) is a value describing relative variation and a wide number of possible

² Ramus et al. (1999) also propose a cluster for mora-timed languages. However, as stated previously the present research only deals with the stress-timed/syllable-timed distinction.

ways to calculate varcos exist. The one used for the present research is calculated as the percentage of the standard variation of the consonantal interval duration (ΔC) of the average duration of consonantal intervals (meanC), i.e.:

$$\text{varco}\Delta C = \frac{\Delta C * 100}{\text{meanC}}; C = \text{duration of consonantal intervals}$$

Dellwo & Wagner (2003) found that there is considerable variation of ΔC on two levels: Within languages syllable rate increases as a function of speakers' intended speech rate (ISR): very slow (s2), slow (s1), normal (no), fast (f1), and fastest possible (f2) (cf. section 2 for a more detailed description and Dellwo et al., forthcoming). At each ISR condition between-language syllable rate is highest for French (F) and lowest for German (G) with English (E) in the middle but closer to German (apart from f2 where G has a higher value than E).

Both between and within languages an increase in syllable rate would always show an increase in ΔC . Supposing a case in which ΔC variation is proportionate to the extent of meanC this would have the following implications on the findings of Dellwo & Wagner (2003): for within- and between-language variation varco ΔC should be rather equal for all intended speech rate versions. This would mean that varco ΔC would not be able to distinguish rhythm classes between languages any more and that ΔC would vary solely as a function of different absolute syllable rates used in the respective languages. It would thus describe syllable rate variation rather than rhythm.

2 Data: The BonnTempo Corpus

A present version of the BonnTempo Corpus (henceforth: BonnTempo) as described in Dellwo et al. (forthcoming) is used for the present research. BonnTempo is a steadily growing database with speakers from different languages assumed to be stress-timed (English, German) and syllable-timed (French, Italian) and not yet uncontroversially classified languages (Polish, Czech). The data in Dellwo & Wagner (2003) use an earlier version of BonnTempo (not called BonnTempo at the time); thus the present research includes their data plus additional speakers.

The data for the present experiments is based on 12 speakers of German (G), 7 of English (E), and 7 of French (F). Speakers were recorded directly on PC in the sound proof booth at the Institute for Communication Research and Phonetics of Bonn University with a high quality condenser microphone (quantisation: 16bit, sampling rate: 44100samples/second). None of the speakers reported any forms of speech or language disorders nor could they be detected during the recording procedure.

A small German text of approximately 80 syllables served as reading material for the German speakers. This text was translated into English and French by philologically educated native speakers of the respective languages. While subjects' voice level was controlled for recording they were asked to get acquainted with the text. After that subjects were recorded reading the text in their native language in what they consider 'normal' reading for this text. Having done that subjects were consecutively asked twice to slow down reading speed, the first time being instructed to read the text 'slowly' and the second time being instructed to read the text 'even slower'. Following this subjects were recorded reading the text in what they consider being 'fast' and then subjects were consecutively asked to increase their reading speed until they reached a version of maximum reading speed according to their opinion or until reading performance became so poor that recordings were terminated. Subjects varied significantly between the number of fast attempts reaching from 3 to 11 fast versions.

Labelling was carried out by human labellers according to phonological syllable durations and consonantal and vocalic intervals on two separate tiers using Praat software (cf. Boersma, 2001). Five versions of each speaker were labelled: 1.) the slowest version (s2), 2.) the slow version (s1), 3.) the

normal version (no), 4.) the fast version (f1), and 5.) the fastest version (f2). From all fast versions the fastest version was decided to be the one that does not show syllable elisions.

Different labellers have carried out labelling work on the data base but inter labeller variation of the same versions has been performed to check inter labeller variability which has generally been regarded as insignificant (cf. Dellwo et al., forthcoming). Automatic labelling software performed poorly especially on the fast versions and was therefore not considered (cf. Steiner, 2004, as well as Dellwo et al., forthcoming).

Furthermore BonnTempo includes recordings of non-native read speech as well as bilinguals for future research on speech rhythm in second languages. New speakers for the current languages are being added and as well as new languages (recordings have been made for Polish and Czech but are not yet labelled). With the appearance of Dellwo et al. (forthcoming) BonnTempo will be made available for free. For further information watch the website of the author (www.phonetiklabor.de).

3 Mean consonantal durations as a function of speech rate

Two things need to be checked first; they may seem obvious but should not be taken for granted: Although Dellwo & Wagner (2003) found a considerable and stable decline of mean syllable duration as a function of speech rate for within and between-language variation it does not necessarily mean that this accounts for consonantal intervals likewise.

For within-language variation it may be that consonantal intervals stay rather stable over the various syllable rates and that vowels play the main compensatory part between the different syllable durations. This may seem rather unlikely for some cases of between-language variation since on this level we are also dealing with systematically varying absolute quantities of consonants: because of the

widely acknowledged fact that French (F) has a simpler syllable structure with less complex consonant clusters than English (E) or German (G) it should be expected that French also shows shorter consonant clusters than E and G and does not only compensate differences in syllable duration on the basis of vowel durations. The case is unclear though for English and German.

In order to investigate whether, and if yes to what extent, within-language variation of syllable rate as a function of intended speech rate (Dellwo & Wagner, 2003) also affects the average durations of consonantal intervals (meanC), meanC was processed for all subjects in BonnTempo at each respective speech rate and then averaged according to F, E, and G. From these values ratios were calculated for s2:s1:no:f1:f2 for each language (F, E, G) with F set to '1' in order to monitor proportional changes and compare them across languages (cf. table 1).

	F	E	G
s2	1	1.34	1.61
s1	1	1.17	1.44
no	1	1.09	1.33
f1	1	1.09	1.30
f2	1	0.93	1.30

Table 2: Ratios for average syllable durations for F:E:G for each intended speech rate (s2, s1, no, f1, f2).

	s2	s1	no	f1	f2
F	1	0.81	0.67	0.60	0.45
E	1	0.93	0.82	0.74	0.66
G	1	0.91	0.81	0.74	0.57

Table 1: MeanC ratios for s2:s1:no:f1:f2 for the three languages E, F, and G with s2 set to 1.

From the values presented in table 1 it can be seen that meanC correlates negatively with intended speech rate. This is true for all languages although proportional changes within languages seem to differ. In this respect G and E show a rather similar proportional decrease of meanC as a function of rate (apart from f2 where G's meanC is smaller than E's), while F's proportional changes are far higher. E.g. meanC at *no* in F is 0.67 times the duration of s2 which is a value that is reached in E only in the fastest possible version (f2).

In analogy to within-language variation, meanC was also monitored for between-language variation in order to find whether languages differing in average syllable duration (Dellwo & Wagner, 2003) also differ in meanC as a function of speech rate. Like in 3.1 ratios for meanC variation were calculated between F, E, and G for each

respective intended speech rate condition (with F set to '1'). Results for the ratios (table 2) reveal that there is a proportional increase in meanC from F to E to G at each respective intended speech rate level (with f2 at E being one exception where there is a decrease of average syllable duration compared to F). From normal to extremely fast speech the proportional changes seem to be rather equal (again apart from f2 at E) while the proportional changes from normal to very slow speech are increasing from F to E to G. In other words: results reveal that mean consonantal intervals are considerably longer in English than they are in French and longest in German.

In conclusion, it has been shown above that the duration of consonantal intervals differ as a function of syllable rate for within- and between-language variation. It may therefore be assumed (cf. introduction) that the standard deviations of consonantal intervals (ΔC) are considerably affected by a higher speech rate, i.e. shorter average durations (meanC). For this reason varco ΔC will be calculated in the next paragraph in order to monitor the proportional ΔC variation according to meanC across different rates for within- and between-language variation.

4 Results for the variation coefficient of ΔC (varco ΔC)

As described previously diagram 1 shows the results for ΔC and %V under the five intended speech rate conditions (s2, s1, no, f1, and f2) for English, German, and French speakers. These results are basically the same as in Dellwo & Wagner (2003) with the only difference that all values have been averaged over a wider range of speakers. It can be seen that the higher number of speakers did not alter the basic finding of Dellwo & Wagner (2003): Generally all languages under investigation show little variation of %V as a function of speech rate (cf. Dellwo & Wagner, 2003, for ratios of %V variation). Regarding ΔC , G and E show **lower** ΔC values than F.

In order to monitor the proportional variation of ΔC as a function of rate a variation coefficient for ΔC (varco ΔC), as explained above, was calculated for all ΔC s at each intended speech rate (s2, s1, no, f1, f2) for all languages (F, E, G). The results plotted against %V can be seen in diagram 2. Results reveal that within languages varco ΔC is differently distributed than ΔC but between languages the general cluster patterns of stress-timed and syllable-timed languages are clearer with varco ΔC than with ΔC since all F versions lie well below E and G on the varco ΔC scale which is not the case for ΔC (cf. diagram 1). In other words: the use of a variation coefficient for ΔC enhances differentiability of rhythm classes for the data presented.

In case of F values for varco ΔC seem to be much less variable than in case of E or G. Another interesting point for F is that varco ΔC correlates positively with speech rate, i.e. proportionally ΔC is higher for shorter intervals than it is for longer intervals. This finding may be seen as support a view that consonantal intervals may not compensate for the reduction of syllable durations as a function of rate in the same way as vowels do since they take proportionately more space in shorter syllables than they do in longer syllables in F.

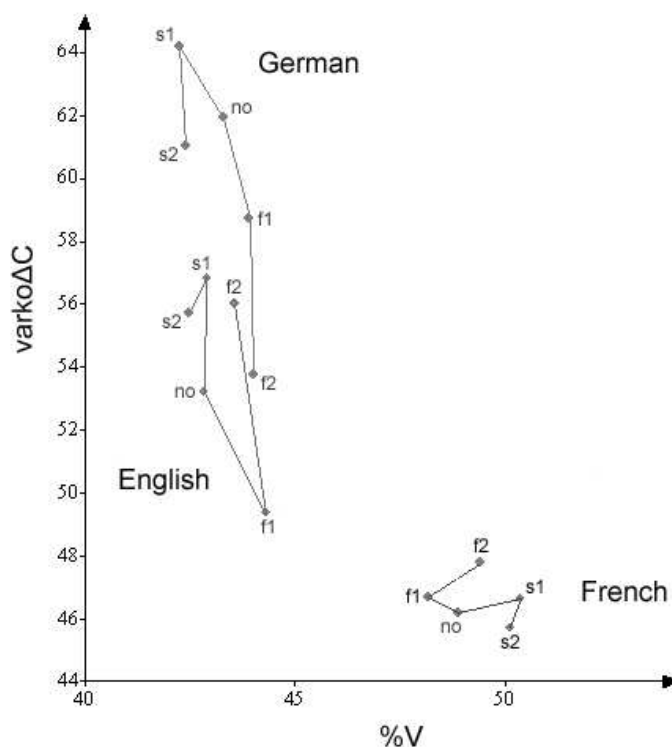


Diagram 2: Results for %V and varco ΔC under different intended speech rate conditions (s2, s1, no, f1, f2) for the languages German, French, and English.

For G and E some values seem to cluster more than others: In case of G the values for s1, s2, and no lie rather close together while varco Δ C values for f1 and f2 decrease strongly with an increase in speech rate. For E, s2 and s1 lie close together while there is a decrease in varco Δ C with increasing speech rate from s1 to no to f1. For f2 varco Δ C rather unexpectedly moves back close to the area of s2, s1.

5 What do the results tell us about the nature of Δ C?

As stated previously the rationale behind the lower Δ C in stress-timed languages is that these languages allow complex consonant clusters while syllable-timed languages do not show this feature (this is most certainly true for the languages in the present study, i.e. French, German, English). It has therefore been stated that values like Δ C describe phonotactic syllable complexity rather than actual rhythmical variations caused by systematic variation of prominent and non-prominent speech units. Considering this, it should be expected that relative changes of Δ C according to speech rate (i.e. varco Δ C) are rather minor since the overall complexity of consonants hardly changes with respect to speech rate. The values for French fulfil this expectation but German and English show that speech rate has a strong effect on the average deviation of consonant cluster duration to their norm. This shows that Δ C is not necessarily determined by absolute syllable complexity but also on the actual realisation of durations of the complex syllable clusters which may vary e.g. as a function of speech rate (as demonstrated in the present case).

6. What do the results tell us about language rhythm?

In case of German, varco Δ C decreases relatively to meanC with increasing speech rate while it approximates values for varco Δ C in the syllable-timed language French. If variations of varco Δ C did represent variations in rhythm then the data presented here would support the following hypothesis in case of French and German: The fact that varco Δ C values for French are rather stable across all speech rates and German values are stable for the slow and normal versions (s2, s1, no) but tend to move towards the French cluster for the fast versions (f1, f2) may lead to the interpretation that German rhythm is not affected by speech rate for normal and slow speech but that it changes towards a syllable-timed rhythm with increasing speech rate. This pattern is not observable for French where rhythm stays rather constant over all possible intended rates. In case of English results are puzzling: English shows a decrease in varco Δ C with increasing speech rate only from s1 to f1, while varco Δ C for f2 moves back closer to s2 and s1. Present plans are to consider a wider range of speakers in the near future to check whether the pattern still holds.

Currently further experiments are under constructing to check whether a change in rhythm for German as a function of speech rate can also be found on a perceptive level. First results from two different observations show support for this theory:

a.) The first sentence of the German text in BonnTempo reads ‘Am nächsten Tag fuhr ich nach Husum’. In terms of syllable prominences nearly all speakers show the following stress pattern for the slow and normal versions: ‘Am nächsten TA_g fU_hr ich nach HU_sum’ (stressed syllables in capitals) while there is a phrase boundary between ‘Tag’ and ‘fuhr’. In the fast versions this pattern seems to change into the following stress pattern: ‘Am nÄ_hsten Tag fU_hr ich nach HU_sum’, while the phrase boundary is dropped. Using nonsense syllables the two stress patterns may be illustrated as following (da = unstressed syllable, **di** = stressed syllable, | = phrase boundary):

normal pattern: da **di** da **di** | **di** da da **di** da
fast pattern: da **di** da da **di** da da **di** da

The Stress clash and the often quoted typical irregular distribution of stressed and unstressed syllables for a stress-timed language like German may make it difficult to hold the normal stress pattern and therefore breaks up into a more regular pattern with a repeating foot (‘**di** da da’). Similar patterns

could be found in the German version of the BonnTempo text while in the French version such a change in stress patterns with increasing rate has not yet been discovered.

So it may be that the change in stress patterns in German stands in connection with the change in values for varco Δ C. Further investigations on this are planned in the future to reveal to what extent stress shifts like the one described actually occur in German and to study the effects they have on Δ C/varco Δ C. If a change in stress pattern should have an influence on varco Δ C and if the regularity of the stress pattern (as in the fast stress pattern) causes a lower varco Δ C then it is assumed that a German speaking style in which all syllables are equally stressed should show the lowest absolute and the lowest variation of varco Δ C according to speech rate. This hypothesis is currently studied in an experiment with German speech in which speakers speak an unnatural rhythm attempting to stress each syllably equally. Results are expected to be reported soon.

b.) In a second attempt currently carried out it is checked whether there is some perceptive evidence for a rhythm change in German on a more general basis (i.e., in sentences that do not show an obvious change of stress pattern as described in a.) for the data in BonnTempo. In this attempt, some normal versions of both German and French are re-synthesised for each speaker to match the overall duration of the fastest version of the respective speaker (all pauses extracted), i.e. an artificial fastest version with the same syllable rate as the respective real fastest version of a speaker is produced. With the same procedure normal versions are re-synthesised from fast version. Native speakers of the respective languages are then presented with the artificial and real versions in contrast and are asked to mark the naturalness of the stimuli. First results from informal experiments reveal that French listeners seem to accept artificial as well as real versions as equally natural while German speakers classify nearly all the re-synthesised versions as unnatural. It will be claimed that this may be proof for a change of German rhythm and stability of French rhythm as function of speech rate. Controlled perception experiments are in progress and results are expected to be reported in the near future.

Acknowledgements

Many thanks go to Ingmar Steiner for helpful comments on the draft version and Petra Wagner for discussions. Furthermore I wish to thank Bianca Aschenberner for her labelling work of BonnTempo, Judith Adrien, Stacy Dellwo, and Franco Ruina for the translation of the experimental material into their native languages (French, English, and Italian), as well as all speakers who contributed to BonnTempo with their voices.

References

- [1] Abercrombie, D. (1967): *Elements of general phonetics*. Aldine: Chicago.
- [2] Barry, W. J., B. Andreeva, M. Russo, S. Dimitrova, and T. Kostadinova (2003): *Do rhythm measures tell us anything about language type?* In: *Proceedings of the 15th ICPHS*, Barcelona, 2693-2696.
- [3] Beckman, M. E. (1992): *Evidence for speech rhythm across languages*. In: Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (eds.) *Speech Perception, Production and Linguistic Structure*. IOS Press: Amsterdam, 457-463.
- [4] Boersma, P. and D. Weenink (2001): *Praat, a system for doing phonetics by computer*. In: *Glott International* (5), 341-345.
- [5] F. Cummins, F. (2002): *Speech rhythm and rhythmic taxonomy*. In *Proceedings of speech prosody*, Aix-en-Provence, 121-136.
- [6] Dellwo, V. and P. Wagner (2003): *Relationships between speech rate and rhythm*. In: *Proceedings of the ICPHS*, pp.
- [7] E. Grabe and E. L. Low (2003): *Durational variability in speech and the rhythm class hypothesis*. In: *Papers in laboratory phonology* (7), 515-546.
- [8] Pike, K. L. (1945): *The intonation of American English*. University Press: Michigan.
- [9] Ramus, F. (2002): *Acoustic correlates of linguistic rhythm: Perspectives*. In: *Proceedings of*

speech prosody, Aix-en-Provence, 115-120.

- [10]Ramus, F., M. Nespó, J. Mehler (1999): *Correlates of linguistic rhythm in the speech signal*. In: *Cognition* (73), 265-292.
- [11]Steiner, I. (in this volume) *A refined acoustic analysis of speech rhythm*.
- [12]Steiner, I. (2004): *Zur Rhythmusanalyse mittels akustischer Parameter*. Magisterarbeit, Bonn, Institut für Kommunikationsforschung und Phonetik.